

METHODS

Observed correlations between person-means depend on within-person correlations

Jonas Haslbeck^{1,2*}, & Sacha Epskamp^{3*}

Received: March 26, 2024 | **Accepted:** August 9, 2024 | **Published:** September 20, 2024 | **Edited by:** Hudson Golino¹Department of Clinical Psychological Science, Maastricht University, ²Department of Psychological Methods, University of Amsterdam, ³Department of Psychology, National University of Singapore

* Both authors contributed equally to this work. Please address correspondence to Jonas Haslbeck, Postbus / P.O. Box 616 6200 MD Maastricht jonashaslbeck@protonmail.com. This article is published under the Creative Commons BY 4.0 license. Users are allowed to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

Intensive longitudinal data have become a common data type across psychological disciplines. A key issue in the analysis of such data is the separation of within-person and between-person effects. This problem is well studied for the effect of between-person effects (e.g., varying intercepts) on within-person parameters (e.g., cross-lagged effects). In this paper, we discuss a less appreciated effect of within-person correlations on correlations between person-wise means. Using simulations and an analytical derivation, we show how observed correlations between person-wise means are a function of both population between-person correlations and within-person correlations. This has implications for the interpretation of statistical relationships between person-wise means, for example when estimated directly from the data or within step-wise approaches to estimating multilevel vector autoregressive models, such as in the popular R package *mIVAR*. We discuss implications for applied research and possible strategies to avoid this problem.

Keywords: longitudinal data analysis; separating within- and between-person effects; multi-level modeling; vector Autoregressive models

1. INTRODUCTION

Cross-sectional data does not allow one to adequately distinguish between within-person effects and between-person effects (Molenaar, 2004). This limitation is one of the driving forces behind the trend towards collecting *intensive longitudinal data* of several subjects measured multiple times (Conner & Barrett, 2012; Hamaker & Wichers, 2017; Kuppens et al., 2022; Miller, 2012; Trull & Ebner-Priemer, 2014). Such data does, in principle, allow for the separation of within- and between-person variance (Epskamp, 2020), which is crucial if the targets of inference are statistical relationships (e.g., correlations, regressions) at the within-person or between-person level. For example, consider a dataset of several subjects measured repeatedly on a set of mood states. A clinical psychologist might be interested in using such data to study within-person relationships to find, for example, relationships between day-to-day fluctuations (states) from the person-wise means. A personality researcher, on the other hand, might be more interested to study between-person relationships to find how the average mood states (traits) relate to one-another.

While multivariate psychometric models exist that adequately separate within- and between-person variance (e.g., Epskamp, 2020; Hamaker et al., 2015; Molenaar, 1985, 2017), they often become intractable to use when the number of variables and/or measurements per person is large. To this end, methods have been proposed to simplify those models. A popular trick is to use the person-wise sample means — the sample mean of all measures of a single person on a particular variable — to stand in for the corresponding true person-wise means. Using the person-wise sample means as a proxy for the true means, the data can then be within-person centered to study within-person effects and relationships between the sample means can be used to study between-person effects (Epskamp et al., 2018; Hamaker & Grasman, 2015). While this method, implemented in the popular R-package *mIVAR* (Epskamp et al., 2018), generally works well for large datasets, it has been recognized that using the person-wise means as such can influence the within-person parameters obtained in an analysis, especially when the number of time points per person is low. For example, a temporal auto-regression can be estimated to be negative due to *Nickell's bias* (Nickell, 1981) when data with too few repeated measures are within-person centered (Burger et al., 2022; Jordan et al., 2020)¹.

While the possibility that between-person effects can bias within-person effects is well known in the literature, here we study the reverse effect. Namely, the possibility that within-person effects are biasing between-person effects. We will show that using the person-wise sample means as proxies for the true person-wise means can lead to bias in estimated *between person parameters*, such as (partial) correlation coefficients between the person-wise means (Epskamp et al., 2022; Epskamp,

¹To understand Nickell's bias, consider a hypothetical longitudinal analysis in which data are within-person centered using sample means and only two observations per person are available. In this case, if the first observation is higher (lower) than the second, the first observation will be above (below) the mean and the second observation will be below (above) the mean, leading to an auto-correlation of -1 . This bias gradually dissipates with more measures per person. For *mIVAR*, a minimum of 20 measures per person is recommended (Burger et al., 2022).

2020). We first illustrate in a simulation study using multilevel vector auto-regressive (VAR) models how correlations between person-wise means can be induced by within-person correlations. That is, we observe correlations between person-wise means, even if the data generating mechanism does not include any such correlation, but only correlations at the within-person level. We then report a second simulation to show that this bias also carries over to the estimation of between-person networks in the context of estimating multilevel VAR models. After illustrating the effect, we provide an analytic derivation of the correlation between person-wise means as a function of population correlations between person-wise means and within-person correlations. We end by discussing implications for interpreting correlation between person-wise means and suggest possible avenues for future research to avoid this bias.

2. WITHIN- AND BETWEEN-PERSON EFFECTS

We begin by introducing notation and terminology. Let $\mathbf{y}_{t,p}$ be a set of n_v variables measured in person $p \in (1, 2, \dots, n_p)$ at time point $t \in \{1, 2, \dots, n_t\}$. We will assume an underlying statistical model that separates $\mathbf{y}_{t,p}$ into a part that is stable for person p , μ_p , and deviation from this stable mean in time point t , $\xi_{t,p}$:

$$\mathbf{y}_{t,p} = \mu_p + \xi_{t,p}. \quad (1)$$

Capitalized subscripts indicate the dimension that is taken to be random. That is, $\mathbf{y}_{T,p}$ indicates responses on a random time point T from fixed person p , and $\mathbf{y}_{t,P}$ indicates responses from a random person P on fixed time-point t . This allows us to establish the terminology about within- and between-person effects which we use in the remainder of this paper; for a more elaborate discussion of this notation, see Epskamp et al. (2022).

2.1 Within-person effects

Within-person effects in observational data typically refer to effects that are generated by (some function of) the variance–covariance structure of a fixed person p within the same window of measurement or across different windows of measurement

$$\text{within-person (co-)variance: } \text{cov}(\mathbf{y}_{T,p}, \mathbf{y}_{T+k,p}) = \text{cov}(\xi_{T,p}, \xi_{T+k,p}),$$

in which $k \in \mathbb{Z}$ (which could be zero) is called the *lag*, indicating a difference across time. This variance is called *within-person* because it is fully derived from within the dataset of person p . When analyzing multiple people, researchers often estimate *fixed-effects*, which describe the within-person variances of the average person, here denoted with the symbol $*$:

$$\text{fixed-effect within-person (co-)variance: } \text{cov}(\mathbf{y}_{T,*}, \mathbf{y}_{T+k,*}) = \text{cov}(\xi_{T,*}, \xi_{T+k,*}).$$

These effects are often simply called *within-person* effects, but it could be argued that this is a misleading label, because the effects are not based on the data of any single subject in the dataset. In this paper, we will circumvent this possible confusion by assuming full homogeneity $\text{cov}(\mathbf{y}_{T,1}, \mathbf{y}_{T+k,1}) = \text{cov}(\mathbf{y}_{T,2}, \mathbf{y}_{T+k,2}) = \dots = \text{cov}(\mathbf{y}_{T,n_p}, \mathbf{y}_{T+k,n_p})$ across persons for simplicity, while noting that any findings do generalize to heterogeneous samples because the fixed-effects adequately describe the covariance structure across people (Epskamp, 2020).

2.2 Between-person effects

Between-person effects and *cross-sectional* effects are often used interchangeably in the literature, but they refer to different variances. *Cross-sectional* effects refer to the variance–covariance structure of cross-sectional data:

$$\text{Cross-sectional variance: } \text{var}(\mathbf{y}_{t,P}),$$

indicating the variance–covariance structure across people on a *fixed time-point* t . This variance–covariance structure, however, is in part also a function of the (fixed-effect) within-person variance covariance structure (see below for more detail). As such, “between-person” is not an adequate term to use for effects established in cross-sectional data, unless it can reasonably be assumed that within-person variance in the data is nil (e.g., when including aggregates over time or stable traits in the data; Epskamp et al., 2018). An adequate definition of *between-person* effects is that these effects rest solely on the variance–covariance structure of the part that is stable within individuals

$$\text{Between-person variance: } \text{var}(\boldsymbol{\mu}_P),$$

that is, the covariances between stable means encoded in $\boldsymbol{\mu}$.

Epskamp et al. (2018) give two examples of situations in which the within-person and between-person correlations between two variables can diverge. The first example, based on Hamaker (2012), discusses a dataset in which authors are measured on the number of spelling errors and typing speed for individual texts; a fixed author might make more spelling errors if they type faster (within-person effect), whereas an author who types fast on average might also make few spelling errors on average (between-person effect). The second example, based on Hoffman (2015), discusses a dataset in which subjects are measured on their physical activity and heart rate; whenever a person is more physically active than average (exercising), their heart rate tends to be higher (within-person effect), and people who often exercise tend to have a lower heart rate on average (between-person effect). In both examples, investigating only one level will not tell the full picture, and performing a cross-sectional analysis will allow only limited conclusions because the results will be a blend of the two sources of variation. This shows that separating within- and between-person effects is often crucial.

2.3 Using person-wise Sample Means as Proxies for True person-wise Means

While modeling techniques exist that adequately aim to separate the within-person and between-person sources of variation (e.g., Asparouhov et al., 2018; Epskamp, 2020), doing so is not always trivial and often leads to computationally heavy estimation procedures that quickly grow intractable for larger numbers of time points and/or variables. To this end, an often-used trick to greatly simplify modeling of such data is to use the person-wise sample means as proxies for the true person-wise means (Hamaker & Grasman, 2015):

$$\hat{\mu}_p = \bar{y}_p = \frac{1}{n_t} \sum_{t=1}^{n_t} y_{t,p} = \frac{1}{n_t} \left(n_t \mu_p + \sum_{t=1}^{n_t} \xi_{t,p} \right) \quad (2)$$

Here, the data can be within-person centered using person-wise sample means \bar{y}_p to obtain within-person effects, and relationships between (or the variance–covariance structure of) elements of \bar{y}_p can be used to obtain estimates of between-person effects. This trick relies on the fact that the person-wise sample means are unbiased maximum likelihood estimators of the population person-wise means. An example of an application of this method is the popular *mIVAR* R package for estimating multilevel graphical vector-autoregressive models (Epskamp et al., 2018), which uses the person-wise sample means to (a) within-person center data for obtaining within-person temporal and contemporaneous effects, and (b) obtain between-person relationships between stable means. We describe the *mIVAR* algorithm in more detail in Appendix A.1.

While the person-wise sample means are indeed unbiased maximum likelihood estimators for the true person-wise means, Equation (2) also shows that $\mu_p \neq \bar{y}_p$ and that the person-wise sample means in \bar{y}_p are in part caused by within-person deviations in $\xi_{t,p}$. This means that the variance–covariance structure of \bar{y}_p will be in part determined by within-person variance - a property that we show formally in Section 4.1. As such, we can expect that analyzing relationships between elements of \bar{y}_p may not only reflect relationships in μ_p , and thus that the estimated between-person relationships based on person-wise sample means may not reflect population between-person relationships. In what follows, we investigate this in more detail. First by illustrating the issue in a simulation study and then by providing mathematical derivation of the variance–covariance structure of \bar{y}_p .

3. ILLUSTRATING DEPENDENCY OF BETWEEN-PERSON EFFECTS ON WITHIN-PERSON CORRELATIONS

We first simulate from a two-variable multilevel VAR model to show how within-person correlations lead to bias in between-person effects as a function of how large the between-person variance is relative to the within-person variance, and sample size. This bias also affects the between-person network estimated with the *mIVAR* method, which we explicitly demonstrate in a second simulation.

3.1 Effects on Between-Person Correlations

We can illustrate the effect of within-person correlations on correlations between person-wise means using any longitudinal model that separates within-person and between-person variances. Here, we generate data under a multilevel lag-1 vector autoregressive (VAR) model with $n_v = 2$ variables, since we suspect most readers will be familiar with it:

$$\mathbf{y}_{t,p} = \boldsymbol{\mu}_p + \mathbf{B}_p (\mathbf{y}_{t-1,p} - \boldsymbol{\mu}_p) + \boldsymbol{\zeta}_{t,p}, \quad (3)$$

where $\boldsymbol{\mu}_p$ are person-specific means, \mathbf{B}_p are person-specific lagged effects of order 1 and $\boldsymbol{\zeta}_{t,p}$ is the person-specific innovation variance. We choose the entries of \mathbf{B}_p such that the VAR model is stable/stationary ([for an accessible introduction to non-stationarity see Ryan et al., 2023](#)).

To keep the illustration simple, we only allowed the means to differ between subjects, and generate homogeneous temporal coefficients in which every element is the same value β :

$$\mathbf{B}_p = \beta \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \forall_p,$$

and set the innovation variance–covariance matrix to an identity matrix to generate standard normal orthogonal innovations:²

$$\text{var}(\boldsymbol{\zeta}_{T,p}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \forall_p.$$

Finally, we generate the true person-wise means for both variables from two independent normal distributions with standard deviation σ_μ :

$$\text{var}(\boldsymbol{\mu}_P) = \sigma_\mu^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

As such, the true between-person covariance/correlation between the two variables equals 0, the parameter β controls the strength of within-person (temporal) relationships between the two variables, and the parameter σ_μ controls the variance in true means across persons. Note that in this example, we manipulate the within-person correlations only with dependencies between variables in lagged effects. However, we could also manipulate within-person correlations by introducing correlations between innovations in $\text{var}(\boldsymbol{\zeta}_{T,p})$. In the simulated example in the following subsection, we will introduce within-person correlations in both ways.

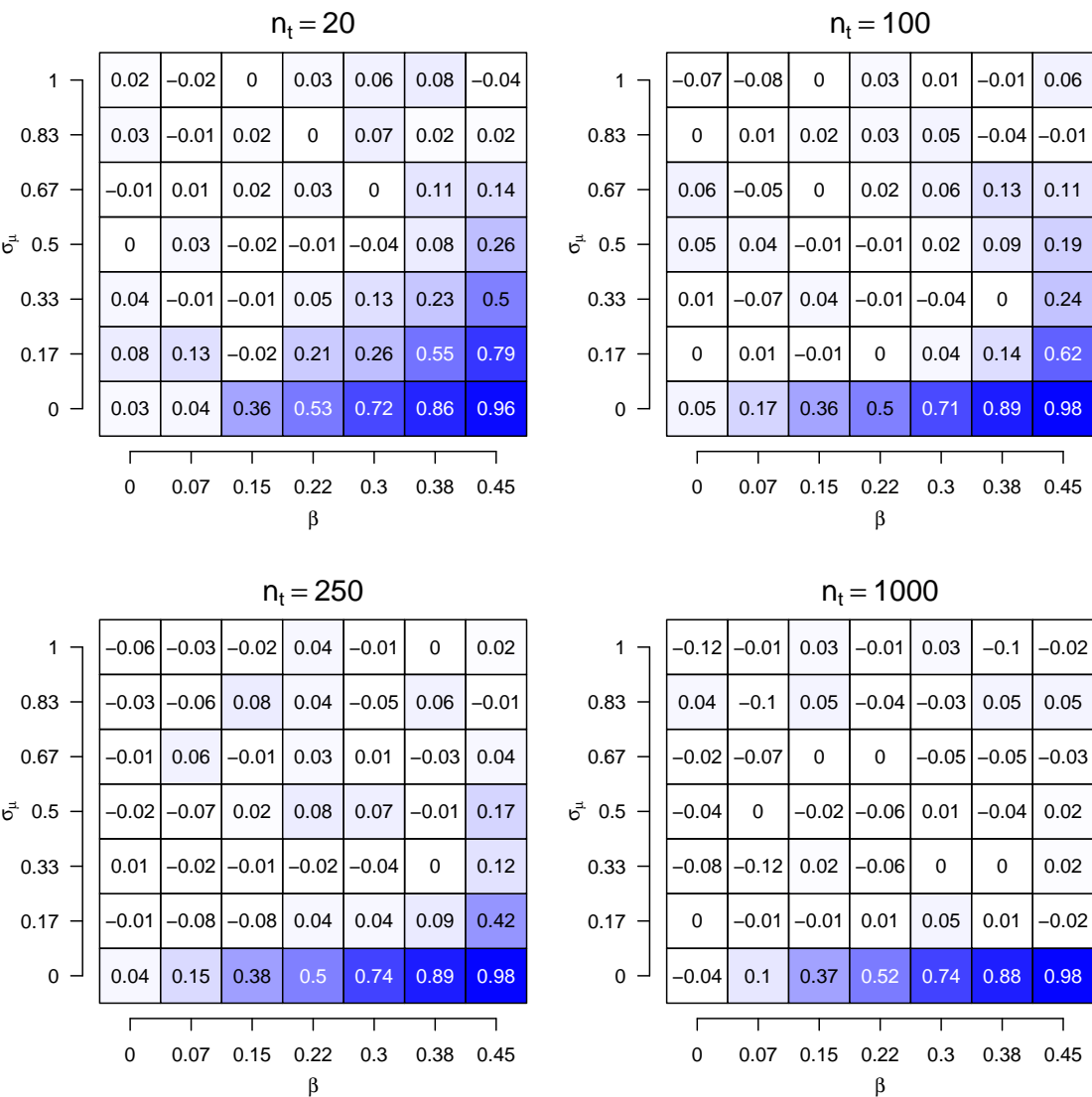
²This also means that we do not simulate data with correlated innovations, which can also be characterized as a *contemporaneous network*. The effects in this contemporaneous network are within-person effects, however, and we expect relationships at the contemporaneous level to similarly impact relationships at the between-person level.

We denote the true between-person correlation with $\rho = \text{cor}(\mu_{1,P}, \mu_{2,P}) = 0$ and the estimated between-person correlation based on the person wise sample means with $\hat{\rho} = \text{cor}(\bar{y}_{1,P}, \bar{y}_{2,P})$. In what follows we simulate data under this multilevel VAR model for different values of β , σ_μ and n_t and evaluate the resulting observed correlation between person-wise means. We varied the within-person strength of relationship β ranging from 0 to 0.45, the between-person strength of relationship σ_μ ranging from 0 to 1, and the number of time points per person across four levels $n_t \in \{20, 100, 250, 1000\}$. We simulate data with a relatively large sample size of $n_p = 500$ subjects in each condition to ensure that the influence of sampling variation on our results is minimal. Given that the only population differences between persons are in the true person-wise means, which are uncorrelated ($\rho = 0$), we might expect that also the observed correlations between sample means ($\hat{\rho}$) will be zero. As such, we are interested in seeing if $\hat{\rho}$ is a good approximation of ρ . Figure 1 shows the estimated correlation between sample means $\hat{\rho}$ rounded to the second decimal mapped to a color gradient from white (0) to blue (1).

Since the true between-person correlation is $\rho = 0$ in the data generating model, any nonzero value for $\hat{\rho}$ in Figure 1 represents bias. If $\beta = 0$ the within-person correlation is equal to zero and therefore also $\hat{\rho} \approx 0$ for all σ_μ and n_t in all simulated condition ($\hat{\rho}$ would be exactly 0 if we let $n_p \rightarrow \infty$). This makes sense, because there is no mechanism in the model that could create a nonzero correlation. However, for $\beta > 0$ we see nonzero correlations between means ($\hat{\rho} > 0$). Specifically, we see that $\hat{\rho}$ increases with β and decreases with higher between-person variation in σ_μ and longer time series n_t . While we will derive these results analytically below, we can already offer the following intuitive explanation: when increasing σ_μ , the proportion of between-person variance in the overall variance term is increasing and as a result the proportion of variance coming from within-person correlations has to decrease. Therefore, the correlation $\hat{\rho}$ has to decrease when increasing σ_μ . For large samples $n_t \rightarrow \infty$, the sample estimates will converge to their true values, which are uncorrelated, which means the covariance has to go to zero. However, this is not necessarily true for normalized measures such as correlations. In this case, $\hat{\rho}$ only converges to zero if $\sigma_\mu = 0$. This can be seen in Figure 1: if $\sigma_\mu > 0$ the correlation $\hat{\rho}$ decreases with increasing n_t ; however, for $\sigma_\mu = 0$ it is independent of n_t .

When is this bias a problem? If we take $\hat{\rho}$ simply as the observed correlation between sample means and draw no conclusion about how this correlation is generated, or acknowledge that it is generated by an unknown combination of within- and between-level effects, then there is no problem. However, if we take the observed correlation between person-wise means $\hat{\rho}$ as an estimate of the population correlation between person-wise means ρ , then we would interpret $\hat{\rho}$ incorrectly, since it is a function of both ρ and within-person correlations. In that sense, $\hat{\rho}$ is a biased estimate of ρ , and Figure 1 shows that this bias can be considerable in conditions in which β is large, σ_μ is small, and n_t is small. This problem extends to other methods which use sample means as proxies of population means, such as the method of R-package *mIVAR* discussed in Section 2.3. We illustrate this next.

Figure 1
The correlation between within-person sample means $\hat{\rho}$, as a function of the size of lagged effects β (x-axis), the true between-person standard deviation σ_{μ} (y-axis), and the length of the time series (panels). Note that if there is no bias, the correlation should be about zero in all cells.



3.2 Effects on Between-Person Network in multilevel VAR Model

The *m*/VAR package uses person-wise sample means as proxies for person-wise population means. Here we illustrate how this approximation leads to biases in the “between-network” consisting of partial correlations between random intercepts in a multilevel VAR model.

To this end, we again use a small simulation. We generated data for $n_p = 250$ subjects on $n_t = 250$ time points on $n_v = 8$ variables. The data generating structures were inspired by previous simulation studies (Du et al., 2024; Epskamp et al., 2018): the temporal network was a skip-1 chain graph



(each variable predictive of itself and a second neighbor) with auto-regressive effects set to 0.5 and cross-lagged coefficients set to 0.25, the contemporaneous network was a chain graph with edge weights (partial correlations) of 0.35, the innovation variances were set to 1, and the between-person network was empty (no between-person covariances). The person-wise means were allowed to vary across people, but no other heterogeneity across persons was included. We generated two datasets: a within-person dataset in which scores for each person were centered on 0, and a set of means for each person with standard deviation of 1. Next, we combined these datasets in four different ways by adding the within-person data to differently scaled simulated means (to control the between-person standard deviation in the generated data). This way, we constructed four datasets: a dataset in which the means were multiplied by 0 to generate a data with no between-person variance ($\sigma_\mu = 0$), and three datasets in which the means were multiplied with different values $\sigma_\mu \in \{0.1, 0.25, 0.5\}$ to generate both within- and between-person variance. We estimated the model parameters using *mIVAR* under default settings with orthogonal random effects.

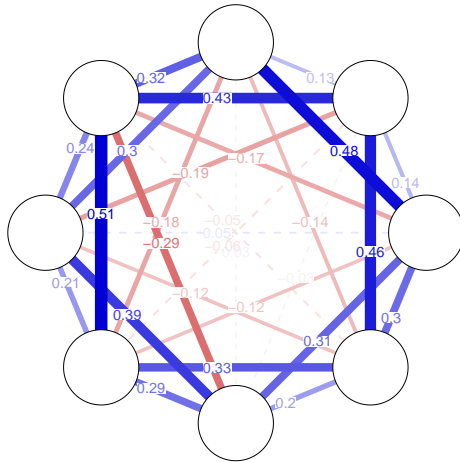
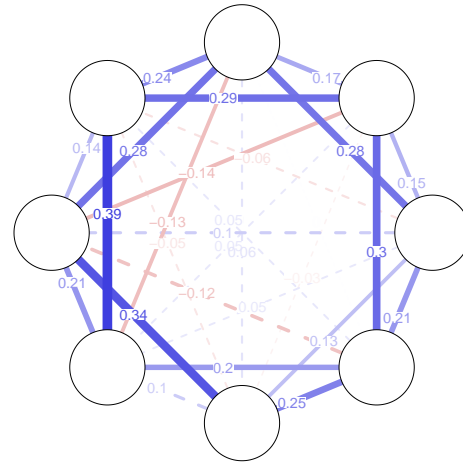
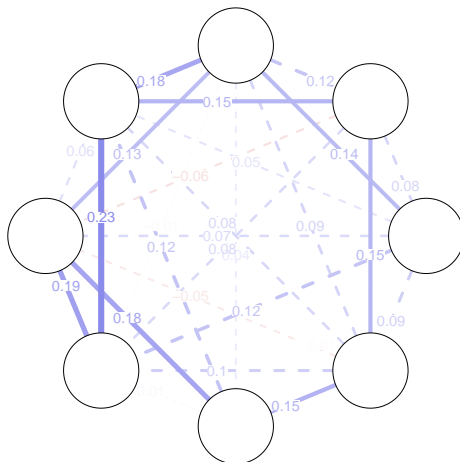
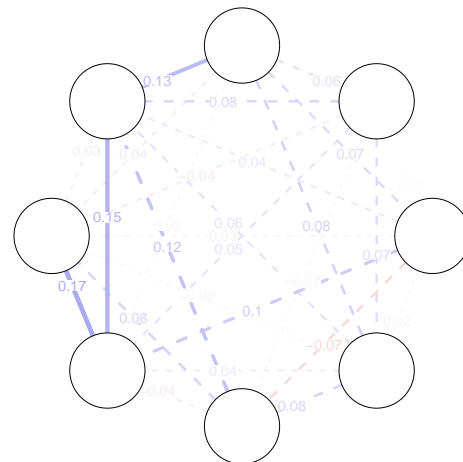
Figure 2 shows the estimated between-person networks. Since the true between-person network is empty, we would hope that the estimates are close to zero and set to zero with probability $1 - \alpha$, where α is the specified significance threshold for hypothesis testing. However, the results show that we estimated strong between-person effects. Most of these effects are not set to zero (solid vs. dashed lines) by a significance test with $\alpha = 0.05$. In fact, we even see negative relationships in the between-person network, despite the fact that we neither have negative relationships in the lagged effects, nor the innovation partial correlation matrix.³ These biases decrease with increasing variance of the true between-person variance, since this implies that the within-person variance component has to become relatively smaller. This is the same effect as illustrated by varying the x-axis in Figure 1 above.

Both illustrations show that when correlating person-wise means, the observed correlations are a function of both within-person and between-person correlations. We can separate the influence of both variance components using hierarchical models. However, if we estimate these models by using person-wise sample means as proxies for their population values, a bias remains. Now that we have illustrated how within-person correlations can bias between-person correlations, we formally derive how the latter are a function of the former.

³However, the model implied within-person lag-0 network does include negative edges, explaining the observed bias in the between-person network. Interestingly, the model-implied *marginal* variance-covariance structure also does not include negative covariances.

Figure 2

Between-person networks estimated with the mIVAR method for between-person standard deviations in means of 0, 0.1, 0.25, and 0.5 on 250 subjects with a time series length of 250 each. Blue edges indicate positive parameters, red edges indicate negative parameters. Solid edges are included at a significance threshold of $\alpha = 0.05$. The data generating (true) between-person network is empty, which means that all shown effects are bias induced by the lagged and contemporaneous within-effects.

A) Estimated between-person network ($\sigma_{\mu} = 0$)B) Estimated between-person network ($\sigma_{\mu} = 0.1$)C) Estimated between-person network ($\sigma_{\mu} = 0.25$)D) Estimated between-person network ($\sigma_{\mu} = 0.5$)

4. DERIVING HOW BETWEEN-PERSON EFFECTS DEPEND ON WITHIN-PERSON EFFECTS

In this section, we will analytically derive the variance-covariance structure of the person-wise sample means in order to investigate how between-person (co-)variances are a function of within-person relationships.

4.1 General Framework

Based on the model in Equation (1), we will assume that the person-wise deviations neither have bias over time nor has bias between people:

$$\mathcal{E}(\xi_{T,p}) = \mathcal{E}(\xi_{t,P}) = \mathbf{0} \quad \forall t,p,$$

in which \mathcal{E} denotes the expected value taken over the capitalized subscript (which indicates the dimension that is random). The expected value of μ will be denoted with a grand mean μ_* :

$$\mu_* = \mathcal{E}(\mu_P).$$

We will assume further that ξ and μ are normally distributed. The variance–covariance matrix of μ can then be interpreted as a *between-person* variance–covariance matrix:

$$\text{Var}(\mu_P) = \Sigma_{\text{bet}}.$$

The covariance of ξ over time can be interpreted as a *within-person* variance–covariance matrix:⁴

$$\text{Cov}(\xi_{i,P}, \xi_{j,P}) = \Sigma_{\text{within}}^{(i,j)}.$$

in which i and j are two different time-points. The assumption that $\mathcal{E}(\xi_{T,p}) = \mathbf{0} \forall p$ naturally implies that $\text{Cov}(\mu_P, \xi_{t,P}) = \mathbf{0} \forall t$. As such, the covariance matrix between any two time points i and j becomes (Epskamp, 2020):

$$\text{Cov}(\mathbf{y}_{i,P}, \mathbf{y}_{j,P}) = \Sigma^{(i,j)} = \Sigma_{\text{bet}} + \Sigma_{\text{within}}^{(i,j)}.$$

Assuming that $\Sigma_{\text{within}}^{(i,j)} \neq \mathbf{0} \forall i,j$, the joint likelihood of all observations can not be simplified beyond modeling the covariance between every item at every time point. To do this, we can form a normally distributed joint vector \mathbf{u} that contains the vectorized time series of person p :

$$\mathbf{u}_p = \begin{bmatrix} \mathbf{y}_{1,p} \\ \mathbf{y}_{2,p} \\ \vdots \\ \mathbf{y}_{n_t,p} \end{bmatrix}, \quad \mathbf{u} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

⁴Epskamp (2020) shows that covariance of ξ across people is the fixed-effect average of the true within-person covariances of ξ across time for a given individual.

with the following distributional parameters:

$$\boldsymbol{\mu} = \mathbf{1}_{n_t} \otimes \boldsymbol{\mu}_* = \begin{bmatrix} \boldsymbol{\mu}_* \\ \boldsymbol{\mu}_* \\ \vdots \\ \boldsymbol{\mu}_* \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}^{(1,1)} & \boldsymbol{\Sigma}^{(1,2)} & \dots & \boldsymbol{\Sigma}^{(1,n_t)} \\ \boldsymbol{\Sigma}^{(2,1)} & \boldsymbol{\Sigma}^{(2,2)} & \dots & \boldsymbol{\Sigma}^{(2,n_t-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}^{(n_t,1)} & \boldsymbol{\Sigma}^{(n_t,2)} & \dots & \boldsymbol{\Sigma}^{(n_t,n_t)} \end{bmatrix},$$

in which $\mathbf{1}_{n_t}$ represents a length n_t vector of ones, and \otimes the Kronecker product.

Of note, while the variance–covariance structure above can be used to understand the behavior of these models, the model parameters are not always estimated through maximum likelihood estimation using the full joint likelihood expressed above. This is especially the case for time-series data where n_t and n_v are large, leading to $\boldsymbol{\Sigma}$ to become an intolerably large matrix that cannot be handled for optimisation purposes by standard commercial computers. To this end, several methods have been proposed that take shortcuts in estimating model parameters from such models, such as analyzing the Topelitz covariance matrix (Hamaker et al., 2002) or by estimating only parts of the model using univariate multilevel regression models (Bringmann et al., 2013).

4.2 Person-wise Sample Means

The person-wise sample means can be obtained as follows. First, we define matrix \boldsymbol{E} as a block matrix that repeats a n_v size identity matrix n_t times:

$$\boldsymbol{E} = \mathbf{1}_{n_t}^\top \otimes \boldsymbol{I} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{I} & \dots & \boldsymbol{I} \end{bmatrix}.$$

Next, we can define the person-wise sample means as follows:

$$\bar{\boldsymbol{y}}_p = \frac{1}{n_t} \boldsymbol{E} \boldsymbol{u}_p.$$

The person-wise sample mean $\bar{\boldsymbol{y}}_p$ is an unbiased estimator of the person-wise mean $\boldsymbol{\mu}_p$, and it follows also that over people the expected value of the person-wise sample mean correctly leads to the grand mean:

$$\mathcal{E}(\bar{\boldsymbol{y}}_P) = \boldsymbol{\mu}_*.$$

The variance of the person-wise sample mean becomes:

$$\begin{aligned}
 \text{Var}(\bar{\mathbf{y}}_P) &= \frac{1}{n_t^2} \mathbf{E} \mathbf{\Sigma} \mathbf{E}^\top \\
 &= \frac{1}{n_t^2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \mathbf{\Sigma}^{(i,j)} \\
 &= \frac{1}{n_t^2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \left[\mathbf{\Sigma}_{\text{bet}} + \mathbf{\Sigma}_{\text{within}}^{(i,j)} \right] \\
 &\neq \mathbf{\Sigma}_{\text{bet}} \quad \text{if } \mathbf{\Sigma}_{\text{within}}^{(i,j)} \neq \mathbf{O} \text{ for any } i \text{ and } j.
 \end{aligned}$$

As such, the variance–covariance structure of the person-wise sample means does *not* necessarily equate to the variance–covariance structure of the true person-wise sample means. Instead, it is a weighted sum, with the $\mathbf{\Sigma}_{\text{bet}}$ matrix weighted n_t^2 times and every possible $\mathbf{\Sigma}_{\text{within}}^{(i,j)}$ matrix weighted once. This means that we cannot expect analyzing relationships between person-wise sample means to lead to correct estimates of population between-person relationships, as within-person relationships can also influence results based on the analysis of person-wise sample means.

If a model is stationary, we can see that $\mathbf{\Sigma}$ takes the form of a Toeplitz matrix with:

$$\begin{array}{ll}
 \mathbf{\Sigma}_{\text{within}}^{\text{lag-0}} = \mathbf{\Sigma}_{\text{within}}^{(1,1)} = \mathbf{\Sigma}_{\text{within}}^{(2,2)} = \dots = \mathbf{\Sigma}_{\text{within}}^{(n_t, n_t)} & \text{included } n_t \text{ times} \\
 \mathbf{\Sigma}_{\text{within}}^{\text{lag-1}} = \mathbf{\Sigma}_{\text{within}}^{(1,2)} = \mathbf{\Sigma}_{\text{within}}^{(2,3)} = \dots = \mathbf{\Sigma}_{\text{within}}^{(n_t-1, n_t)} & \text{included } 2(n_t - 1) \text{ times} \\
 \mathbf{\Sigma}_{\text{within}}^{\text{lag-2}} = \mathbf{\Sigma}_{\text{within}}^{(1,3)} = \mathbf{\Sigma}_{\text{within}}^{(2,4)} = \dots = \mathbf{\Sigma}_{\text{within}}^{(n_t-2, n_t)} & \text{included } 2(n_t - 2) \text{ times} \\
 \vdots & \\
 \mathbf{\Sigma}_{\text{within}}^{\text{lag}-(n_t-1)} = \mathbf{\Sigma}_{\text{within}}^{(1, n_t)} & \text{included 2 times.}
 \end{array}$$

We can reasonably assume also then that the covariance between observations far away in time will be negligible small:

$$\mathbf{\Sigma}_{\text{within}}^{\text{lag}-k} \rightarrow \mathbf{O} \quad \text{for higher values of } k,$$

as otherwise the time-series would not be stationary. As such, we can derive the following implications:

- If $\mathbf{\Sigma}_{\text{bet}} = \mathbf{O}$, all effects found between person-wise sample means are due to within-person covariances (or Type 1 errors).
- These effects will typically be small, but can lead to large standardized parameters (e.g., partial correlation coefficients).
- If $\mathbf{\Sigma}_{\text{bet}} \neq \mathbf{O}$, we expect that $\mathbf{\Sigma}_{\text{bet}}$ will dominate $\text{Var}(\bar{\mathbf{y}}_P)$ for increasing time-series lengths n_t .

As such, we expect that analyzing relationships between sample means will lead to accurate inference on true between-person effects for increasing time-series lengths n_t and pending on the strength of the between-person variance in the data generating mechanism. If there is no between-person variance, relationships between person-wise sample means will reflect within-person relationships regardless of the sample size (n_p) or length of time-series (n_t). Note that the number of subjects n_p does not impact this bias. However, we can note that larger sample sizes n_p do reduce the standard errors on the estimated relationships between sample means. This means that we can expect that increasing the sample size n_p but retaining a low number of repeated measures per person n_t might lead to an increase in Type 1 errors in estimated between-person relationships. The accuracy of the estimates of person-wise means can only be improved by increasing the time-series length per person. These expectations are also visible in Figure 1, which shows that the bias is worse when the between-person variance is lower, the time-series length is shorter and the within-person relationship is stronger.

5. DISCUSSION

We showed that when analyzing a longitudinal dataset consisting of multiple people measured on multiple time points, the commonly used method of using the person-wise sample means as proxies for true person-wise means in order to uncover between person effects can be biased by within-person effects. This has implications for the interpretation of models for the statistical relationships between person-wise means, and for methods that make use of person person-wise sample means, such as the popular R package *mIVAR* for multilevel graphical vector auto-regressive (VAR) models.

The main implication of these results is that when analyzing statistical relationships between person-wise means by directly modeling them or as part of the multilevel VAR model in *mIVAR*, these relationships cannot be interpreted as pure between-person effects. Instead, the correlation will be a function of both the population-between person effect and within-person correlations, and which contributes to which degree is unknown when analyzing empirical data. We found that the bias will decrease as a function of the time-series lengths and the proportion of between-person variance, but also that situations are conceivable where the bias is considerably large in reasonable between-person variances and time-series lengths. Our results are related to those of Hamaker (2023) who showed that correlations between cross-sections of multiple time series (i.e., a single time point of each measure of each subject, instead of the within-person sample analyzed here) is also a function of both within-person and between-person effects.

One way in which this bias can be avoided is by jointly estimating the model parameters corresponding to within-person and between-person effects in a single step. For the multilevel VAR model, this can be done in two accessible software suites: the R package *psychometrics* (Epskamp, 2020) and the Dynamic Structural Equation Modeling (DSEM) module in *MPlus* version 8 and higher (Asparouhov

et al., 2018; McNeish & Hamaker, 2020). Both methods, however, also have practical downsides. The implementation of the model in *psychometrics* only allows for estimating random means (no random network structures) and is only suitable for panel data in which the number of time points is relatively low (e.g., 3-10) and the number of variables is moderate (up to about 20-30 variables). The DSEM model in MPlus, on the other hand, is not freely available and that currently it is computationally feasible only for 5-6 variables. MPlus also does not allow the direct modeling of partial correlation structures. An important avenue for future work would therefore be to provide a free implementation of DSEM that scales to larger number of variables. Another possibility could be to still estimate the multilevel VAR model in two steps such as in the *mIVAR* package, but to iterate between the two steps until convergence. Studying such an approach could be promising, however, it will also have to deal with the considerable computational cost involved in estimating many multilevel models.

6. ACKNOWLEDGEMENTS

We would like to thank Oisín Ryan and Lourens Waldorp for helpful discussions about the topic of this paper. We would like to thank Janne Adolf for carefully reviewing our paper and our reproducibility archive, and spotting a coding error in an earlier version of this manuscript. JMBH has been supported by the gravitation project ‘New Science of Mental Disorders’ (www.nsmdeu), supported by the Dutch Research Council and the Dutch Ministry of Education, Culture and Science (NWO gravitation grant number 024.004.016). SE was supported by National University of Singapore grant A-8001098-00-00.

7. MATERIALS

Code to reproduce the simulation shown in Figure 1 can be found at <https://github.com/jmbh/person-wiseMeanCor>.

8. CONFLICTS OF INTEREST

The authors report no competing interests.

REFERENCES

- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 359–388. <https://doi.org/10.1080/10705511.2017.1406803>
- Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PloS one*, 8(4), e60188. <https://doi.org/10.1371/journal.pone.0060188>
- Burger, J., Hoekstra, R. H. A., Mansueto, A. C., & Epskamp, S. (2022). Network estimation from

time series and panel data. In *Network psychometrics with R* (pp. 169–192). Routledge. <https://doi.org/10.4324/9781003111238-13>

- Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic medicine*, 74(4), 327–337. <https://doi.org/10.1097/PSY.0b013e3182546f18>
- Du, X., Skjerdingsstad, N., Freichjel, R., Ebrahimi, O. V., Hoekstra, R. H. A., & Epskamp, S. (2024). Moving from exploratory to confirmatory network analysis: An evaluation of sem fitindices and cutoff values in network psychometrics. *PsyArXiv preprint*. <https://doi.org/10.31234/osf.io/d76ab>
- Epskamp, S., Hoekstra, H. A., Burger, J., & Waldorp, L. J. (2022). Chapter 9: Longitudinal Design Choices: Relating Data to Analysis. In A. M. Isvoranu, S. Epskamp, L. J. Waldorp, & D. Borsboom (Eds.), *Network Psychometrics with R: A guide for behavioral and social scientists*. Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9781003111238-12>
- Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*, 85(1), 206–231. <https://doi.org/10.1007/s11336-020-09697-3>
- Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, 82, 904–927. <https://doi.org/10.1007/s11336-017-9557-x>
- Epskamp, S., Waldorp, L. J., Möttus, R., & Borsboom, D. (2018). The gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, 53(4), 453–480. <https://doi.org/10.1080/00273171.2018.1454823>
- Hamaker, E. L. (2012). Why researchers should think “within-person”: A paradigmatic rationale. In *Handbook of research methods for studying daily life* (pp. 43–61). The Guilford Press New York, NY.
- Hamaker, E. (2023). The curious case of the cross-sectional correlation. *Multivariate Behavioral Research*, 1–12. <https://doi.org/10.1080/00273171.2022.2155930>
- Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. (2002). On the nature of sem estimates of arma parameters. *Structural Equation Modeling*, 9(3), 347–368. https://doi.org/10.1207/S15328007SEM0903_3
- Hamaker, E. L., & Grasman, R. P. (2015). To center or not to center? investigating inertia with a multilevel autoregressive model. *Frontiers in psychology*, 5, 1492. <https://doi.org/10.3389/fpsyg.2014.01492>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological methods*, 20(1), 102–116. <https://doi.org/10.1037/a0038889>
- Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, 26(1), 10–15. <https://doi.org/10.1177/0963721416666518>
- Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge. <https://doi.org/10.4324/9781315744094>
- Jordan, D. G., Winer, E. S., & Salem, T. (2020). The current status of temporal network analysis for clinical science: Considerations as the

paradigm shifts? *Journal of Clinical Psychology*, 76(9), 1591–1612. <https://doi.org/10.1002/jclp.22957>

- Kuppens, P., Dejonckheere, E., Kalokerinos, E. K., & Koval, P. (2022). Some recommendations on the use of daily life methods in affective science. *Affective Science*, 3(2), 505–515. <https://doi.org/10.1007/s42761-022-00101-0>
- Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Clarendon Press.
- McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in mplus. *Psychological methods*, 25(5), 610–635. <https://doi.org/10.1037/met0000250>
- Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on psychological science*, 7(3), 221–237. <https://doi.org/10.1177/1745691612441215>
- Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: Interdisciplinary Research & Perspective*, 2(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1
- Molenaar, P. C. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50(2), 181–202. <https://doi.org/10.1007/BF02294246>
- Molenaar, P. C. (2017). Equivalent dynamic models. *Multivariate behavioral research*, 52(2), 242–258. <https://doi.org/10.1080/00273171.2016.1277681>
- Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the econometric society*, 49(6), 1417–1426. <https://doi.org/10.2307/1911408>
- Ryan, O., Haslbeck, J., & Waldorp, L. (2023). Non-stationarity in time-series analysis: Modeling stochastic and deterministic trends. *PsyArXiv preprint*. <https://doi.org/10.31234/osf.io/z7ja2>
- Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current directions in psychological science*, 23(6), 466–470. <https://doi.org/10.1177/0963721414550706>

A. APPENDIX

A.1 The mIVAR algorithm: two-step multilevel graphical vector autoregression

A popular method in which we expect this effect to occur is the *two-step multilevel graphical vector auto-regression* algorithm implemented in the *mIVAR* package for R (Epskamp et al., 2018). We will term this model the *mIVAR-model* and the corresponding estimation algorithm the *mIVAR-algorithm* for the remainder of this paper. The *mIVAR-algorithm* aims to estimate parameters corresponding to a multilevel lag-1 graphical vector auto-regressive model:

$$\mathbf{y}_{t,p} = \boldsymbol{\mu}_p + \mathbf{B}_p (\mathbf{y}_{t-1,p} - \boldsymbol{\mu}_p) + \boldsymbol{\zeta}_{t,p}.$$

Here, the \mathbf{B}_p matrix is a person-specific matrix of temporal coefficients, which is typically used to draw a *temporal network* structure. The variance–covariance structure of the innovation term $\boldsymbol{\zeta}$ can be modeled as a Gaussian graphical model (GGM; Lauritzen, 1996)—a model of partial correlation coefficients (Epskamp et al., 2017):

$$\text{Var}(\boldsymbol{\zeta}_{T,p}) = \boldsymbol{\Sigma}_p^{(\zeta)} = \boldsymbol{\Delta}_p^{(\zeta)} \left(\mathbf{I} - \boldsymbol{\Omega}_p^{(\zeta)} \right)^{-1} \boldsymbol{\Delta}_p^{(\zeta)},$$

in which $\boldsymbol{\Omega}_p$ contains partial correlation coefficients—typically visualized as a person-specific *contemporaneous network*—and $\boldsymbol{\Delta}_p$ is a diagonal scaling matrix. When taking the person subscript as random, as needed for the sampling distribution of the person-specific sample mean, we can replace these matrices with their fixed-effects counterparts (Epskamp, 2020):

$$\text{Var}(\boldsymbol{\zeta}_{\dagger,P}) = \boldsymbol{\Sigma}_*^{(\zeta)} = \boldsymbol{\Delta}_*^{(\zeta)} \left(\mathbf{I} - \boldsymbol{\Omega}_*^{(\zeta)} \right)^{-1} \boldsymbol{\Delta}_*^{(\zeta)}, \quad (4)$$

in which the \dagger symbol stands for ‘any time-point’ and the asterisk subscript $*$ denotes the fixed-effect (average over people). The person-specific mean can also be modeled as a GGM:

$$\boldsymbol{\Sigma}_{\text{bet}} = \text{Var}(\boldsymbol{\mu}_P) = \boldsymbol{\Sigma}^{(\mu)} = \boldsymbol{\Delta}^{(\mu)} \left(\mathbf{I} - \boldsymbol{\Omega}^{(\mu)} \right)^{-1} \boldsymbol{\Delta}^{(\mu)}. \quad (5)$$

Epskamp (2020) show that the implied within-person fixed-effect variance–covariance matrices of this model take the following forms:

$$\begin{aligned} \text{vec} \left(\boldsymbol{\Sigma}_{\text{within}}^{\text{lag}-0} \right) &= (\mathbf{I} \otimes \mathbf{I} - \mathbf{B}_* \otimes \mathbf{B}_*) \text{vec} \left(\boldsymbol{\Sigma}_*^{(\zeta)} \right) \\ \boldsymbol{\Sigma}_{\text{within}}^{\text{lag}-k} &= \mathbf{B}_* \boldsymbol{\Sigma}_{\text{within}}^{\text{lag}-(k-1)} \quad \text{if } k \neq 0 \end{aligned} \quad (6)$$

Estimating this model directly through multivariate optimization, however, quickly grows prohibitively intractable for increasing values of n_t and n_v .

Epskamp et al. (2018) propose to estimate the parameters of the *mIVAR-model* through a series of univariate multilevel regressions, by combining the work of Bringmann et al. (2013) on estimating multilevel temporal networks with the work of Hamaker and Grasman (2015) on separating within- and between-person effects. The *mIVAR-algorithm* comprises of two steps, each consisting of n_v multilevel models to be estimated:

- Step 1: Temporal and between-person effects are estimated by using each variable once as dependent variable, predicted by all within-person centered lagged variables (for the temporal effects) and person-wise sample means of each other variable (for the between-person effects).
- Step 2: The estimated residuals from the Step 1 models are used in a new sequence of univariate multilevel models in which each residual is once treated as dependent variable predicted by the remaining residuals.

Clearly, the Step 1 models heavily rely on person-wise sample means to within-person center the lagged predictors as well as to estimate the between-person effects. We therefore expect the bias we demonstrate above to occur in the *mIVAR-algorithm*. Below, we study in simulations the extent of this bias.